Query complexity in expectation

Jedrzej Kaniewski* Troy Lee† Ronald de Wolf‡

Abstract

We study the query complexity of computing a function $f : \{0,1\}^n \to \mathbb{R}_+$ *in expectation*. This requires the algorithm on input x to output a nonnegative random variable whose expectation equals $f(x)$, using as few queries to the input x as possible. We exactly characterize both the randomized and the quantum query complexity by two polynomial degrees, the nonnegative literal degree and the sum-of-squares degree, respectively. We observe that the quantum complexity can be unboundedly smaller than the classical complexity for some functions, but can be at most polynomially smaller for functions with range $\{0,1\}$.

These query complexities relate to (and are motivated by) the extension complexity of polytopes. The *linear* extension complexity of a polytope is characterized by the randomized *communication* complexity of computing its slack matrix in expectation, and the *semidefinite* (psd) extension complexity is characterized by the analogous quantum model. Since query complexity can be used to upper bound communication complexity of related functions, we can derive some upper bounds on psd extension complexity by constructing efficient quantum query algorithms. As an example we give an exponentially-close entrywise approximation of the slack matrix of the perfect matching polytope with psd-rank only $2^{n^{1/2+\varepsilon}}$. Finally, we show there is a precise sense in which randomized/quantum query complexity in expectation corresponds to the Sherali-Adams and Lasserre hierarchies, respectively.

*Centre for Quantum Technologies, National University of Singapore and QuTech, Delft University of Technology, the Netherlands. j.kaniewski@nus.edu.sg.

†School of Mathematics and Physical Sciences, Nanyang Technological University and Centre for Quantum Technologies, Singapore. troyjlee@gmail.com. This material is based on research supported in part by the Singapore National Research Foundation under NRF RF Award No. NRF-NRFF2013-13.

‡CWI and University of Amsterdam, the Netherlands. rdewolf@cwi.nl. Partially supported by a Vidi grant from the Netherlands Organization for Scientific Research (NWO) which ended in 2013, by ERC Consolidator Grant QPROGRESS, and by the European Commission IST STREP project Quantum Algorithms (QALGO) 600700.

1 Introduction

1.1 Computing functions in expectation

We study the complexity of computing a function $f : \{0,1\}^n \to \mathbb{R}_+$ *in expectation*. In this setting, on input x we want our algorithm to output a nonnegative real number whose expectation (over the algorithm's internal randomness) exactly equals $f(x)$. Getting the expectation right is an easier task than computing the function value $f(x)$ itself, and suffices in some applications. For example, suppose we want to approximate the value $F(x) = \sum_{i=1}^m f_i(x)$ that depends on $x \in \{0,1\}^n$. Then we can just compute each $f_i(x)$ *in expectation* and output the sum of the results. By linearity of expectation, the output will have expectation $F(x)$, and it will be tightly concentrated around its expectation if the random variables are not too wild (so the Central Limit Theorem applies). It is not necessary to compute or even approximate any of the values $f_i(x)$ themselves for this. This illustrates that computing functions in expectation is an interesting model in its own right. Additionally, it is motivated by connections with the *extension complexity* of polytopes that are used in combinatorial optimization (roughly: the minimal size of linear or semidefinite programs for optimizing over such a polytope), as we describe below in Section 1.2.

The complexity of computing f can be measured in different ways, and here we will focus on *query* complexity. We measure the complexity of computing a function in expectation by the (worst-case) number of queries to the input $x \in \{0,1\}^n$ that the best algorithm uses. We study both *randomized* and *quantum* versions of this model and show that both of these query complexities can be exactly characterized by natural notions of polynomial degree. In Section 3 we show that the randomized query complexity of computing f in expectation equals the "nonnegative literal degree" of f, which is the minimal d such that f can be written as a nonnegative linear combination of products of up to d variables or negations of variables. In Section 4 we show that the quantum complexity equals the "sum-of-squares degree", which is the minimal d such that there exist polynomials p_i of degree at most d satisfying $f(x) = \sum_i p_i(x)^2$ for all $x \in \{0,1\}^n$.

In Section 5 we observe that quantum and classical query complexities (equivalently: the above two types of polynomial degree) can be arbitrarily far apart. For example, the function $f(x) = (\sum_{i=1}^n x_i - 1)^2$ is the square of a degree-1 polynomial and hence can be computed in expectation with only 1 quantum query, while randomized algorithms need n queries to get this expectation right. In contrast, we also show that for functions with range $\{0,1\}$, the quantum-classical gap cannot be very large: at most cubic.

Lower bounds on the quantum query complexity can be obtained from lower bounding the sum-of-squares degree of the function at hand, which is often non-trivial. In Section 6, using techniques from approximation theory, we prove that the function $f(x) = (\sum_{i=1}^n x_i - 1)(\sum_{i=1}^n x_i - 2)$ has sum-of-squares degree $\Omega(\sqrt{n})$. Hence quantum algorithms require $\Omega(\sqrt{n})$ queries to compute this function in expectation.

1.2 Motivation: linear and semidefinite extension complexity

Our main motivation for studying query complexity in expectation comes from combinatorial optimization, in particular from linear and semidefinite programs. Many optimization problems can be formulated as maximizing or minimizing a linear function over a polytope. For example, in the Traveling Salesman Problem on n-vertex undirected graphs, one wants to minimize a linear function (the length of the tour) over the polytope $P \subseteq \mathbb{R}^{\binom{n}{2}}$ that is the convex hull of all Hamiltonian cycles in the complete n-vertex graph K_n. If this polytope could somehow be represented as the feasible region of a small linear or semidefinite program, then we could efficiently solve the problem using the ellipsoid or interior-point methods.

Informally, the **linear extension complexity** of a polytope $P \subseteq \mathbb{R}^d$ is the minimum number of linear inequalities (over the d variables of P as well as possibly auxiliary variables) whose feasible region projects

down to P. If the linear extension complexity is small, there is a small linear program to optimize over P.

Motivated by erroneous claims [Swa86] that the TSP polytope had polynomial linear extension complexity (implying P = NP), Yannakakis [Yan91] showed that "symmetric" linear extensions of the Traveling Salesman Polytope need $2^{\Omega(n)}$ linear inequalities. He showed the same for the perfect matching polytope (which is spanned by all perfect matchings in K_n), despite the fact that finding a maximum matching can be done efficiently! For a long time, generalizing these lower bounds to arbitrary (possibly non-symmetric) linear extensions was an open question. However, recently Fiorini et al. [FMP$^+$12] proved a $2^{\Omega(n^{1/2})}$ lower bound on the linear extension complexity of the TSP polytope. Subsequently Rothvoß [Rot14] proved a $2^{\Omega(n)}$ lower bound for the perfect matching polytope, which via a reduction implies the same bound for TSP. Chan et al. [CLRS13] obtained lower bounds on linear extension complexity for constraint satisfaction problems via a different route: roughly put, they showed that arbitrary linear extensions are not much more powerful than the specific linear extensions produced by the "Sherali-Adams Hierarchy"; hence they could obtain lower bounds on linear extension complexity from known bounds on the Sherali-Adams hierarchy.

The **positive semidefinite (psd) extension complexity** of polytope P is similar, but replaces the linear programs by potentially more powerful semidefinite programs. The complexity is now the minimal dimension of a semidefinite program whose feasible region projects down to P. In contrast to the case of linear extension complexity, very few lower bounds on psd extension complexity are known. Until recently, there were only a few lower bounds for "symmetric" psd extensions [LRST14, FSP13]. However, in a *very* recent breakthrough, Lee et al. [LRS14] generalized the approach of [CLRS13] to show that arbitrary psd extensions are not much more powerful than the specific psd extensions produced by the "Lasserre Hierarchy". In particular they showed that the TSP polytope has psd extension complexity $2^{\Omega(n^{1/13})}$.

Surprisingly, there is a very close connection between these extension complexities and the model of computing functions in expectation, albeit for the *communication complexity* of computing a 2-input function. More precisely, suppose Alice receives input x, Bob receives input y, and they want to compute some function $g(x, y)$ (which may also be viewed as a matrix). In the usual setting of communication complexity [KN97], one of the parties (let's say Bob) has to output this value $g(x, y)$ exactly, either with probability 1 or with high probability. However, we may also consider how much communication they need to compute $g(x, y)$ *in expectation*, i.e., now Bob needs to output a nonnegative random variable whose expected value equals $g(x, y)$. Faenza et al. [YFGT12] showed that the logarithm of the linear extension complexity of a polytope P equals the randomized communication complexity of computing (in expectation) a matrix associated with P, known as the *slack matrix*. Lifting this result to the quantum/psd case, Fiorini et al. [FMP$^+$12] showed that the logarithm of the *psd* extension complexity equals the one-way *quantum* communication complexity of computing the slack matrix of P in expectation; in this model Alice sends a single quantum message to Bob. These connections show that studying (linear and psd) extension complexity of a polytope P is *equivalent* to studying (randomized and one-way quantum) communication complexity in expectation, of the slack matrix of P.

How do our results on the *query* complexity of computing a function in expectation impact this *communication* complexity? Many functions of interest in communication complexity are of the form $g(x, y) = f(x \wedge y)$ for some Boolean function $f : \{0, 1\}^n \to \{0, 1\}$, where the AND-connective is applied bitwise. Functions of this form also arise as (submatrices of) slack matrices of interesting polytopes, for example the correlation polytope. Quite generally across the usual models of worst-case complexity, be it deterministic, randomized, or quantum, upper bounds on the *query complexity* of f imply upper bounds on the *communication complexity* of g. In Section 7 we show that this also holds for the randomized and quantum models of computing a function in expectation. As this leads to multi-round communication protocols, we also show that the one-way quantum communication complexity of computing a function in expectation equals the

two-way complexity.

In Section 7.3 we give an application of the connection between query algorithms and communication complexity (equivalently, *psd rank*), by deriving an exponentially-close entrywise approximation of the slack matrix S of the perfect matching polytope with psd rank $2^{n^{1/2+\varepsilon}}$. This psd rank is surprisingly low in view of the fact that Rothvoß [Rot14] showed that the nonnegative rank of S is $2^{\Omega(n)}$, and Braun and Pokutta [BP15] showed that any $\tilde{S}$ that is $O(1/n)$-close to S still needs nonnegative rank $2^{\Omega(n)}$.

Communication protocols derived from query algorithms have a specific structure. In spirit, this is somewhat similar to looking at linear/psd extensions derived from hierarchies of specific linear or semidefinite programs like the Sherali-Adams and Lasserre hierarchies. We show that these two relaxations actually correspond in a precise sense: just as the linear and psd extension complexities are characterized by models of communication complexity in expectation, the Sherali-Adams and Lasserre hierarchies are characterized by randomized and quantum models of query complexity in expectation, respectively. This connection, described in Section 2.4, follows from known characterizations of these hierarchies in terms of notions of polynomial degrees which exactly correspond to the polynomial degrees we consider here.

2 Preliminaries

2.1 Polytopes and extension complexity

A polytope $P \subseteq \mathbb{R}^d$ has both an *inner description* as the convex hull of a set $V \subseteq \mathbb{R}^d$ of points, $P = \mathrm{conv}(V)$; and an *outer description* as the intersection of halfspaces, $P = \{x \in \mathbb{R}^d : Ax \leq b\}$. A *slack matrix* integrates information from these two descriptions:

Definition 1 *Let $P = \mathrm{conv}(V) = \{x : Ax \leq b\}$ be a polytope. The slack matrix M of P has columns labeled by $v \in V$ and rows labeled by constraints $A_i x \leq b_i$, with entries $M(i, v) = b_i - A_i v$.*

Definition 2 *Let M be a nonnegative matrix. A nonnegative factorization of M of size d consists of two sets of d-dimensional nonnegative vectors $\{a_x\}, \{b_y\}$ such that $M(x, y) = a_x^T b_y$ for all x, y. The nonnegative rank of M, denoted $\mathrm{rk}_+(M)$, is the minimal size among all nonnegative factorizations of M. Equivalently, it is the minimum number of nonnegative rank-one matrices whose sum is M.*

Definition 3 *Let M be a nonnegative matrix. A psd factorization of M of size d consists of two sets of d-by-d psd matrices $\{A_x\}, \{B_y\}$ such that $M(x, y) = \mathrm{Tr}(A_x B_y)$ for all x, y. The psd rank of M, denoted $\mathrm{rk}_{\mathrm{psd}}(M)$, is the minimal size among all psd factorizations of M.*

Note that a nonnegative factorization is a psd factorization where the matrices are diagonal.

The *linear extension complexity* of a polytope P is the minimum number of facets of a (higher-dimensional) polytope which projects to P. The *semidefinite (psd) extension complexity* of P is the minimum d such that an affine slice of the cone of d-by-d positive semidefinite matrices projects to P. These complexity measures can be captured in terms of the above notions of rank of a slack matrix:

Theorem 4 ([Yan91, GPT13]) *The linear extension complexity of a polytope P is the nonnegative rank of a slack matrix of P. The semidefinite (psd) extension complexity of P is the psd rank of a slack matrix of P.*

A polytope may have different slack matrices associated with it, depending on which inner and outer description are used. By Theorem 4 these slack matrices all have the same nonnegative and psd rank.

2.2 Candidate matrix for lower bounding the correlation polytope

One of our targets is the correlation polytope: $\mathrm{COR}_n = \{xx^T : x \in \{0,1\}^n\}$. Fiorini et al. [FMP$^+$12] showed that lower bounds on the linear/semidefinite extension complexity of the correlation polytope imply lower bounds on several other polytopes of interest, including the Traveling Salesman Polytope. The next lemma from [Pad89] gives a family of matrices that occur as a submatrix of the slack matrix of the correlation polytope.

Lemma 5 *Let $p(z) = a + bz + cz^2$ be a single-variate degree-2 polynomial that is nonnegative on nonnegative integers. The matrix $M(x,y) = p(|x \wedge y|)$ for $(x,y) \in \{0,1\}^n$ is a submatrix of a slack matrix for the correlation polytope COR_n.*

Proof: As p is nonnegative on nonnegative integers, $-bz - cz^2 \leq a$ is a valid inequality for integers $z \geq 0$. Note that $\mathrm{Tr}(xx^T yy^T) = |x \wedge y|^2$ and $\mathrm{Tr}(\mathrm{diag}(x) yy^T) = |x \wedge y|$ for all $x, y \in \{0,1\}^n$. Thus $\mathrm{Tr}((-b \cdot \mathrm{diag}(x) - c \cdot xx^T) yy^T) \leq a$ is a valid inequality, whose slack is $p(|x \wedge y|)$. Note that the columns of M are labeled by vertices of the correlation polytope yy^T for $y \in \{0,1\}^n$ and likewise the constraints are labeled by xx^T for $x \in \{0,1\}^n$. □

Later in this paper we will consider the matrix $M(x,y) = (|x \wedge y| - 1)(|x \wedge y| - 2)$ and its associated query problem $f(x) = (|x| - 1)(|x| - 2)$, where $|x|$ denotes the Hamming weight of the Boolean string x.

2.3 Polynomials

We will study two types of polynomials that are obviously nonnegative on the Boolean cube: nonnegative literal polynomials and sum-of-squares polynomials.

Definition 6 (nonnegative literal degree) *A nonnegative literal polynomial is a nonnegative linear combination of products of variables and negations of variables, i.e., it can be written as*

$$p(x) = \sum_{S \subseteq [n]} \sum_{b \in \{0,1\}^{|S|}} \alpha_{S,b} \prod_{i \in S} ((-1)^{b_i} x_i + b_i)$$

where each $\alpha_{S,b} \geq 0$. Its degree is $\max\{|S| : \alpha_{S,b} \neq 0\}$. The nonnegative literal degree of $f : \{0,1\}^n \to \mathbb{R}_+$, denoted $\mathrm{ldeg}_+(f)$, is the minimum degree of a nonnegative literal polynomial p that equals f on $\{0,1\}^n$.

Such p are also called *nonnegative juntas* [CLRS13].

Definition 7 (sum-of-squares degree) *Let d be a natural number. A sum-of-squares polynomial of degree d is a polynomial p that can be written in the form*

$$p(x) = \sum_{i \in \mathcal{P}} p_i(x)^2,$$

where $\mathcal{P}$ is a finite index set and the p_i are polynomials of degree $\leq d$. The sum-of-squares (sos) degree of $f : \{0,1\}^n \to \mathbb{R}_+$, denoted $\deg_{sos}(f)$, is the minimum d for which such a p equals f on $\{0,1\}^n$.

Note that a sum-of-squares polynomial of degree d is actually a polynomial of degree $2d$; we allow this slight abuse of notation in order to give a clean characterization in Theorem 12 below.

2.4 The Sherali-Adams and Lasserre hierarchies

Consider the optimization problem

$$\alpha(f) = \max_{x \in \{0,1\}^n} f(x) \tag{1}$$

where f is given by a multilinear polynomial. Many important optimization problems can be cast in this framework, including NP-hard ones. For example finding the maximum cut in a graph $G = (V, E)$ with n vertices corresponds to the quadratic function $f(x) = \sum_{(i,j) \in E} x_i(1 - x_j)$.

If $c \geq \alpha(f)$, then the function $c - f$ is nonnegative on $\{0,1\}^n$. One way we can witness this is by expressing $c - f$ as a polynomial which is obviously nonnegative for all $x \in \{0,1\}^n$. The *Sherali-Adams hierarchy* [SA90] looks for a witness in the form of a nonnegative literal polynomial. The sum-of-squares or *Lasserre hierarchy* looks for a witness in the form of a sum-of-squares polynomial [Las01, Par00, Sho87].

If we can find a nonnegative literal polynomial p of degree d such that $c - f(x) = p(x)$, then this witnesses that the optimal value is upper bounded as $\alpha(f) \leq c$. Moreover, determining if the nonnegative literal polynomial degree of $c - f(x)$ is at most d can be formulated as a linear program of size $n^{O(d)}$. The value of the d-round Sherali-Adams relaxation for (1) is the smallest value of c such that $c - f(x)$ is a degree-d nonnegative literal polynomial. Thus the smallest d for which a Sherali-Adams relaxation certifies an *optimal* upper bound is exactly the nonnegative literal degree $\mathrm{ldeg}_+(\alpha(f) - f)$ of the function $\alpha(f) - f$.

Similarly, if we can find polynomials $p_i : \{0,1\}^n \to \mathbb{R}$ of degree at most d, such that $c - f(x) = \sum_i p_i(x)^2$, then this witnesses that $\alpha(f) \leq c$. Moreover, searching for such polynomials p_i can be expressed as a semidefinite program of size $n^{O(d)}$. The smallest value of c such that $c - f$ is degree-d sum-of-squares is known to be equivalent to the relaxation of (1) given by the d^{th} level of the Lasserre hierarchy. The level of the Lasserre hierarchy required to exactly capture (1) is thus $\deg_{sos}(\alpha(f) - f)$.

3 Randomized query complexity in expectation

In this section we study classical randomized query complexity in expectation, characterize it by the nonnegative literal degree, and relate it to the Sherali-Adams hierarchy.

3.1 Definition

We define a randomized model of computing a function in expectation. A *randomized decision tree* is a probability distribution μ over deterministic decision trees. We consider deterministic decision trees with leaves labeled by nonnegative real numbers. A randomized decision tree computes a function $f : \{0,1\}^n \to \mathbb{R}_+$ if for every $x \in \{0,1\}^n$ the expected output of the tree on input x is $f(x)$. The *cost* of such a tree is, as usual, the maximum cost, that is the length of a longest path from the root to a leaf, of a deterministic decision tree that has nonzero μ-probability.

Definition 8 *The randomized query complexity of computing f in expectation, denoted $\mathrm{RE}(f)$, is the minimum cost among all randomized decision trees that compute f in expectation.*

3.2 Characterization of $\mathrm{RE}(f)$ by polynomials

We now show that $\mathrm{RE}(f)$ is characterized by the nonnegative literal degree.

Theorem 9 *Let $f : \{0,1\}^n \to \mathbb{R}_+$. Then $\mathrm{RE}(f) = \mathrm{ldeg}_+(f)$.*

Proof: $\mathrm{RE}(f) \geq \mathrm{ldeg}_+(f)$. We need to show how a randomized decision tree induces a nonnegative literal polynomial. First consider a deterministic decision tree T with leaves labeled by nonnegative real numbers. For each path p from root to leaf, we construct a literal monomial m_p where x_i appears in m_p if $x_i = 1$ is on p, and $1 - x_i$ appears if $x_i = 0$ is on p. The coefficient α_p of m_p is the label of the leaf of p. If we let $q_T(x) = \sum_{\text{paths } p} \alpha_p m_p(x)$ then we have that $q_T(x)$ is equal to the output of the tree on input x. Moreover the degree of q_T is at most the depth of T. Now for a randomized decision tree that chooses a deterministic decision tree T with probability $\mu(T)$, we set the polynomial $r(x) = \sum_T \mu(T) q_T(x)$, which gives a nonnegative literal representation of f.

$\mathrm{RE}(f) \leq \mathrm{ldeg}_+(f)$. Let

$$p(x) = \sum_{S \subseteq [n]} \sum_{b \in \{0,1\}^{|S|}} \alpha_{S,b} \prod_{i \in S} ((-1)^{b_i} x_i + b_i)$$

be a nonnegative literal polynomial representing f of degree $\mathrm{ldeg}_+(f)$. Let $M = \sum_{S,b} \alpha_{S,b}$. The algorithm chooses S, b with probability $\alpha_{S,b}/M$ and query all $i \in S$ to evaluate $a_{S,b} = \prod_{i \in S} ((-1)^{b_i} x_i + b_i)$. Output $M \cdot a_{S,b}$. The expected output on input x equals $p(x)$, and the number of queries is $\leq \mathrm{ldeg}_+(f)$. □

Referring back to Section 2.4, this gives a connection between randomized query complexity in expectation and the Sherali-Adams hierarchy: the smallest d for which a Sherali-Adams relaxation certifies the optimal upper bound $\alpha(f)$ on the maximization problem (1), is exactly $\mathrm{RE}(\alpha(f) - f)$.

4 Quantum query complexity in expectation

Here we study *quantum* query complexity in expectation, characterize it by sum-of-squares degree, and relate it to Lasserre. We assume familiarity with quantum computing [NC00] and query complexity [BW02].

4.1 Definition

We define the quantum query complexity of computing a function $f : \{0,1\}^n \to \mathbb{R}_+$ in expectation. A T-query algorithm is described by unitaries $U_0, \ldots, U_T$ and a POVM $\{E_\theta\}_{\theta \in \Theta}$, where each E_θ is a psd matrix labeled by nonnegative real θ, and $\sum_{\theta \in \Theta} E_\theta = I$. As usual, on input x the query algorithm proceeds from the initial state $|\bar{0}\rangle$ by alternately applying a unitary and the query oracle O_x (which maps $|i, b\rangle \mapsto |i, b \oplus x_i\rangle$), so that the state of the algorithm after t queries is $|\psi_x^t\rangle = U_t O_x \ldots O_x U_1 O_x U_0 |\bar{0}\rangle$. Let $E = \sum_{\theta \in \Theta} \theta E_\theta$. As the probability of output θ upon measuring $|\psi_x^T\rangle$ is $\mathrm{Tr}(E_\theta |\psi_x^T\rangle\langle\psi_x^T|)$, the expected value of the output is $\mathrm{Tr}(E|\psi_x^T\rangle\langle\psi_x^T|)$. The algorithm *computes f in expectation* if $f(x) = \mathrm{Tr}(E|\psi_x^T\rangle\langle\psi_x^T|)$ for every $x \in \{0,1\}^n$.

Definition 10 *The quantum query complexity of computing f in expectation, denoted $\mathrm{QE}(f)$, is the minimum T for which there is a T-query quantum algorithm computing f in expectation.*

4.2 Characterization of $\mathrm{QE}(f)$ by polynomials

We now adapt the polynomial method [BBC+01] to characterize $\mathrm{QE}(f)$. The key is the following lemma, which says that the amplitudes of the final state of a T-query algorithm are degree-T polynomials in x:

Lemma 11 ([BBC+01]) *The state $|\psi_x^t\rangle$ of a quantum query algorithm on input x after t queries can be written as $\sum_{i,z} \alpha_{i,z}(x)|i, z\rangle$, where each $\alpha_{i,z}(x)$ is an n-variate multilinear polynomial in x of degree $\leq t$.*

Theorem 12 *Let $f : \{0,1\}^n \to \mathbb{R}_+$. Then* $\mathrm{QE}(f) = \deg_{sos}(f)$.

Proof: $\underline{\mathrm{QE}(f) \geq \deg_{sos}(f).}$ Say there is a T-query algorithm to compute f in expectation. Then

$$f(x) = \sum_\theta \theta \langle \psi_x^T | E_\theta | \psi_x^T \rangle.$$

As the coefficients θ are nonnegative real numbers, it suffices to show that each term $\langle \psi_x^T | E_\theta | \psi_x^T \rangle$ can be written as the sum of squares of polynomials of degree at most T.

Let $E_\theta = \sum_i \lambda_i |e_\theta^i\rangle\langle e_\theta^i|$ be the eigenvalue decomposition of E_θ, where each $\lambda_i \geq 0$. Then

$$\langle \psi_x^T | E_\theta | \psi_x^T \rangle = \sum_i \lambda_i |\langle \psi_x^T | e_\theta^i \rangle|^2 \ .$$

We have that $\langle \psi_x^T | e_\theta^i \rangle$ is a linear combination of amplitudes of $|\psi_x^T\rangle$, hence by Lemma 11 it is a degree $\leq T$ polynomial in x. Since $\lambda_i \geq 0$ this gives a representation of $\langle \psi_x^T | E_\theta | \psi_x^T \rangle$ as a sum-of-squares polynomial of degree $\leq T$. Hence $T \geq \deg_{sos}(f)$.

$\underline{\mathrm{QE}(f) \leq \deg_{sos}(f).}$ Let $d = \deg_{sos}(f)$. We first exhibit a quantum algorithm for the special case where $f = p^2$ for some degree-d polynomial p. This is inspired by the proof of [Wol03, Theorem 2.3]. Let $p = \sum_s \widehat{p}(s)(-1)^{x \cdot s}$ be the Fourier representation of p, where s ranges over $\{0,1\}^n$. Because p has degree d, we have $\widehat{p}(s) \neq 0$ only if $|s| \leq d$. The algorithm is as follows:

1. Prepare n-qubit state $c \sum_s \widehat{p}(s)|s\rangle$, where $c = 1/\sqrt{\sum_s \widehat{p}(s)^2}$ is a normalizing constant.
2. Apply a unitary that maps $|s\rangle \mapsto (-1)^{x \cdot s}|s\rangle$ for all s of weight $|s| \leq d$; one can show that this can be implemented using d queries.
3. Apply the n-qubit Hadamard transform to the state.
4. Measure the state and output $2^n/c^2$ if the measurement result was 0^n, otherwise output 0.

Note that the amplitude of the basis state $|0^n\rangle$ after step 3 is

$$\frac{c}{\sqrt{2^n}} \sum_s \widehat{p}(s)(-1)^{x \cdot s} = \frac{c}{\sqrt{2^n}} p(x).$$

Hence the probability that the final measurement results in outcome 0^n is $(\frac{c}{\sqrt{2^n}} p(x))^2$, and the expected value of the output is $(\frac{c}{\sqrt{2^n}} p(x))^2 \cdot 2^n/c^2 = p(x)^2 = f(x)$, as desired.

Now consider the general case where $f = \sum_{i \in \mathcal{P}} p_i^2$. The algorithm chooses one $i \in \mathcal{P}$ uniformly at random and runs the above algorithm to produce an output with expected value $p_i(x)^2$. It finally outputs that output multiplied by $|\mathcal{P}|$. Clearly, the algorithm uses at most d queries to x, and the expected value of its final output is

$$\frac{1}{|\mathcal{P}|} \sum_i p_i(x)^2 |\mathcal{P}| = \sum_i p_i(x)^2 = f(x).$$

Hence $\mathrm{QE}(f) \leq d = \deg_{sos}(f)$. □

This gives a surprising connection between quantum query complexity in expectation and the Lasserre hierarchy: the smallest level d of the Lasserre hierarchy that certifies the optimal upper bound $\alpha(f)$ on the maximization problem (1), is exactly $\mathrm{QE}(\alpha(f) - f)$.

5 Gaps and relations between $\mathrm{RE}(f)$ and $\mathrm{QE}(f)$

For some $f : \{0,1\}^n \to \mathbb{R}_+$, the quantum query complexity in expectation $\mathrm{QE}(f)$ can be *much* smaller than its classical counterpart $\mathrm{RE}(f)$. An extreme example is the n-bit function $f(x) = (|x| - 1)^2$, where $\mathrm{QE}(f) = 1$ by Theorem 12, but $\mathrm{RE}(f) = n$. The latter holds because on the all-0 input the algorithm needs to produce a nonzero output with positive probability, but on weight-1 inputs it can never output anything nonzero, hence a classical algorithm needs n queries on the all-0 input.

In contrast, if the range of f is Boolean, then $\mathrm{QE}(f)$ is at most polynomially smaller than $\mathrm{RE}(f)$:

Theorem 13 *For every $f : \{0,1\}^n \to \{0,1\}$ we have $\mathrm{RE}(f) \leq 16\mathrm{QE}(f)^3$.*

Proof: The result follows by chaining the following three inequalities:

1. $\mathrm{RE}(f)$ is obviously at most the deterministic decision tree complexity of f, denoted $D(f)$;
2. $D(f) \leq 2\deg(f)^3$ by a result of Midrijanis [Mid04, Theorem 4];
3. $\deg(f) \leq 2\mathrm{QE}(f)$, because by Theorem 12 a T-query QE-algorithm gives a degree-T sum-of-squares polynomial that represents f, which is a polynomial of degree $\leq 2T$. □

The main reason this query complexity result is interesting is that the analogous statement for *communication* complexity is equivalent to the longstanding log-rank conjecture! The communication version of Theorem 13 would say that for all *Boolean* matrices M, the quantum and classical communication complexity of computing M in expectation are at most polynomially far apart. As noted by Fiorini et al. [FMP+12], this is equivalent to $\log \mathrm{rk}_+(M) \leq \mathrm{polylog}(\mathrm{rk}_{\mathrm{psd}}(M))$, which in turn is equivalent to the log-rank conjecture. Presumably such a communication version will be substantially harder to prove than the above query version. However, in many cases results in query complexity "mirror" (often much harder) results in communication complexity, so our Theorem 13 may be viewed as (weak) evidence for the log-rank conjecture.

6 Quantum query complexity lower bound

In this section we show that the function $f(x) = (|x| - 1)(|x| - 2)$ has $\mathrm{QE}(f) = \Omega(\sqrt{n})$. We do this by showing the corresponding lower bound on the sum-of-squares degree of f, adapting techniques from approximation theory commonly used to show quantum query lower bounds in the bounded-error model.

We do this by using Theorem 12 and bounding the sum-of-squares degree. As is common in query complexity lower bounds by the polynomial method [BBC+01], we will use a symmetrization argument to define a single-variate polynomial $Q : \mathbb{R} \to \mathbb{R}$ that behaves well on $[n]$, and then use Markov's lemma from approximation theory to bound the degree of Q.

A new complication in our setting is the following. If $f(x) = \sum_i p_i(x)^2$ then we would like to define a "symmetrized" polynomial $g : [n] \to \mathbb{R}$ where $g(k) = \mathbb{E}_{x:|x|=k}\left[\sum_i p_i(x)^2\right]$. We do not know how to show, however, that g remains a nonnegative polynomial. To get around this, we define symmetrized polynomials $q_i(k) = \mathbb{E}_{x:|x|=k}[p_i(x)]$ for each p_i individually, then recombine the symmetrized polynomials as $Q(k) = \sum_i q_i(k)^2$. We are then able to bound the sum-of-squares degree of Q.

Theorem 14 *Let $f(x) = (|x| - 1)(|x| - 2)$ for $x \in \{0,1\}^n$. Then $\deg_{sos}(f) \geq \sqrt{n/48}$.*

Proof: Suppose that f can be expressed as

$$f(x) = \sum_i p_i(x)^2,$$

where $\deg(p_i) \leq T$ for all i. Let $q_i : [n] \to \mathbb{R}$ be defined as $q_i(k) = \mathbb{E}_{|x|=k}[p_i(x)]$. By a standard symmetrization argument [MP87], each q_i is a polynomial of degree at most T. Now consider

$$Q(k) = \sum_i q_i(k)^2,$$

which is a nonnegative polynomial in k of degree at most $2T$. It satisfies $Q(0) = 2$, since there is only one x of weight 0. Also, $Q(1) = Q(2) = 0$ since $f(x) = 0$ for $|x| \in \{1, 2\}$. The zeroes of a nonnegative polynomial must have even multiplicity, so at least 2. Therefore there must exist a polynomial q of degree at most $2T - 4$ such that

$$Q(k) = (k-1)^2(k-2)^2 q(k).$$

By convexity of the quadratic function, we find that

$$Q(k) = \sum_i |q_i(k)|^2 = \sum_i \left|\mathbb{E}_{|x|=k}[p_i(x)]\right|^2 \leq \sum_i \mathbb{E}_{|x|=k}\left[|p_i(x)|^2\right] = \mathbb{E}_{|x|=k}[f(x)] = (k-1)(k-2),$$

which implies

$$q(k) \leq 1/(k-1)(k-2). \tag{2}$$

Note that $q(k) \leq 1/6$ for all integers $k \in \{4, \ldots, n\}$.[1] We now simply lower bound the degree of q using the following lemma of Markov:

Lemma 15 (Markov) *If q is a real polynomial then* $\deg(q) \geq \sqrt{\frac{n}{2} \cdot \frac{\max_{x\in[0,n]} |q'(x)|}{\max_{x\in[0,n]} |q(x)|}}$.

Here q' denotes the derivative of q. Since $q(0) = Q(0)/4 = 1/2$, we know that the maximum value of q in the interval $[0, n]$ is at least $1/2$. Now suppose $\max_{x\in[0,n]} |q(x)| = c \geq 1/2$, and say that this maximum is attained at x^*. Since $q(k) \leq 1/6$ for all integers $k \in \{4, \ldots, n\}$, we know x^* is at most distance 4 from an x where $q(x) \leq 1/6$. Thus $|q'(x)| \geq (6c-1)/24$ for some $x \in [0, n]$. This, together with $c \geq 1/2$, shows that the ratio in Markov's lemma is at least

$$\frac{6c-1}{24c} = \frac{1}{4} - \frac{1}{24c} \geq \frac{1}{6}.$$

Thus overall we obtain $2T \geq \deg(q) \geq \sqrt{\frac{n}{2}\frac{1}{6}} = \sqrt{n/12}$, implying the lower bound. □

We note that stronger lower bounds on sum-of-squares degree are known for related functions. Let $k = \lfloor \frac{n}{2} \rfloor$ and consider $g(x) = (x_1 + \cdots + x_n - k)(x_1 + \cdots + x_n - k - 1)$. This polynomial is nonnegative on all $x \in \{0, 1\}^n$, and the induced matrix $M_g(x, y) = g(x \wedge y)$ is a submatrix of the slack matrix of the correlation polytope by Lemma 5. For odd n, Grigoriev [Gri01] shows that the sum-of-squares degree of g is $\lfloor \frac{n}{2} \rfloor$ (see also [Lau03]). Blekherman et al. [BGP14] show that g even has high *rational* sum-of-squares degree: if the product pg has sos degree d, where p is an sos polynomials of degree r, then $r + d \geq \lfloor \frac{n}{2} \rfloor$.

[1] While we know that q is nonnegative on $[n]$, we will not use this information.

Our lower bound technique is quite different from those used in these works, and is more closely related to works showing bounds on the minimum degree of a polynomial that *approximates* a function in ℓ_∞ norm. In fact, our proof has recently been extended by Arunachalam, Yuen, and the last author [AYW14] to show that this $\Omega(\sqrt{n})$ sos-degree lower bound remains valid for functions g that approximate f pointwise up to additive error $O(1/n)$. This is important because the very recent framework of Lee et al. [LRS14] uses lower bounds on the sum-of-squares degree of a function that approximates f pointwise to show lower bounds on the psd rank of a matrix associated with f.

7 Psd rank and query complexity in expectation

7.1 Psd rank characterizes two-way quantum communication complexity

Fiorini et al. [FMP$^+$12] defined a *one-way* model of quantum communication to compute a matrix in expectation, and showed that this complexity is characterized by the logarithm of the psd rank. We show below that this characterization continues to hold for the more general *two-way* communication model, which allows multiple rounds of communication between the two parties Alice and Bob. Hence one-way and two-way quantum communication complexity are the same for computation in expectation.

We will not formally define the model of two-way quantum communication complexity (see [Wol02] for more technical details), instead just highlighting the differences of the model of computing a function in expectation to the normal model. As usual, Alice and Bob each start with their own input, x and y respectively, and then the protocol specifies whose turn it is to speak and what message they send to the other party. At the end of the protocol Bob must output a *nonnegative* number, which is a random variable z that depends on the inputs x and y as well as on the internal randomness of the protocol.

The major difference with the usual model is the notion of when a protocol is correct. Let M be a matrix with nonnegative real entries whose rows are indexed by Alice's possible inputs, and whose columns are indexed by Bob's inputs. We say a protocol *computes the matrix M in expectation* if, for every (x, y), $M(x, y)$ equals the expected value of the output z on input (x, y). As usual, the *cost* of the protocol is the worst-case number of qubits that are communicated (summed over all rounds).

Definition 16 *The quantum communication complexity of computing a matrix M in expectation, denoted* $\mathrm{QCE}(M)$*, is the minimum q such that there exists a quantum protocol of cost q that computes M in expectation. The minimum q when we restrict to one-way protocols is denoted* $\mathrm{QCE}^1(M)$.

The following theorem shows that two-way quantum communication complexity is not more powerful than its one-way cousin: both are characterized by the psd rank.

Theorem 17 $\log \mathrm{rk}_{\mathrm{psd}}(M) \leq \mathrm{QCE}(M) \leq \mathrm{QCE}^1(f) \leq \log(\mathrm{rk}_{\mathrm{psd}}(M) + 1)$.

Proof: The second inequality is obvious from the definitions. Fiorini et al. [FMP$^+$12] already showed how to construct a one-way protocol to compute M in expectation using $\log(\mathrm{rk}_{\mathrm{psd}}(M) + 1)$ many qubits, establishing the third inequality. Thus we focus on the first inequality. Given a general protocol that computes M in expectation using q qubits of communication, we need to construct a psd factorization of size 2^q.

The first step is to observe that one can replace the range of outputs of a multi-round q-qubit QCE-protocol by $\{0, m\}$, where m is the maximum output among all runs of the protocol: instead of outputting m', just output m with probability m'/m and 0 with probability $1 - m'/m$, which preserves the expected value of the output. In the remainder of the proof we assume for ease of notation that $m = 1$.

Now use the Kremer-Yao lemma [Kre95, Yao93] on this modified multi-round q-qubit communication protocol: its final state on input x, y can be written as

$$\sum_{i \in \{0,1\}^{q+1}} |a_i(x)\rangle |i_{q+1}\rangle |b_i(y)\rangle,$$

where $|a_i(x)\rangle$ and $|b_i(y)\rangle$ are non-normalized states, and i_{q+1} is the last bit of string i, corresponding to the output (0 or 1). Define 2^q-by-2^q psd matrices $A_x(i,j) = \langle a_i(x)|a_j(x)\rangle$ where i, j range over all $(q+1)$-bit strings that end in 1. Similarly define B_y. The expected value of the output is the probability to output 1:

$$\| \sum_{i \in \{0,1\}^q \times \{1\}} |a_i(x)\rangle |1\rangle |b_i(y)\rangle \|^2 = \sum_{i,j \in \{0,1\}^q \times \{1\}} \langle a_i(x)|a_j(x)\rangle \cdot \langle b_i(y)|b_j(y)\rangle = \mathrm{Tr}(A_x B_y).$$

Thus a multi-round q-qubit protocol gives a psd factorization of M of size 2^q. □

7.2 Upper bounds on psd rank from quantum algorithms

We now show that efficient quantum query algorithms for computing functions $f : \{0,1\}^n \to \mathbb{R}_+$ in expectation give rise to an efficient quantum communication protocol to compute the matrix $M_f(x,y) = f(x \wedge y)$ in expectation, and hence to a low-rank psd factorization of M_f. We state it more generally:

Theorem 18 *Let Y be a finite set. For every $y \in Y$, let $f_y : \{0,1\}^n \to \mathbb{R}_+$ satisfy $\mathrm{QE}(f_y) \leq T$. Define a $2^n \times |Y|$ matrix M by $M(x,y) = f_y(x)$. Then $\mathrm{QCE}(M) \leq 2T(\log(n)+1)$, and hence $\mathrm{rk}_{\mathrm{psd}}(M) \leq (2n)^{2T}$.*

Proof: The proof is very similar to an analogous statement by Buhrman, Cleve, and Wigderson [BCW98] for regular quantum communication complexity. Bob (who has input y) runs a T-query algorithm for f_y; whenever he needs to make a query to x he sends the $(\log(n)+1)$-qubit query register to Alice, who applies the query and sends it back. Thus every query is implemented using $2(\log(n)+1)$ qubits of communication, and the expected value of Bob's output is $f_y(x)$. The bound on the psd rank follows from Theorem 17. □

Lee et al. [LRS14] independently proved a similar upper bound on psd rank, stated in terms of the sos degree of the f_y rather than quantum query complexity (which are equal by Theorem 12).

The $\log n$ factor in Theorem 18 is necessary. Consider the function $f(x) = (|x|-1)^2$. Then $\mathrm{QE}(f) = 1$ by Theorem 12. On the other hand $\mathrm{rk}_{\mathrm{psd}}(M_f) \geq n/\sqrt{2}$: it is easy to see that the rank of M is at most the square of its psd rank, and the rank of $M_f(x,y) = (|x \wedge y| - 1)^2$ is $n^2/2 + 1$ using [BW01, Section 4.1].

7.3 Application: approximating the slack matrix of the matching polytope

Here we give an application of the above connection between query algorithms and psd rank, by deriving an exponentially-close entrywise approximation of the slack matrix S of the perfect matching polytope, by a matrix with psd rank not much bigger than $2^{\sqrt{n}}$. This shows a big difference to the case of nonnegative rank: Braun and Pokutta [BP15] show that any $\tilde{S}$ that is $O(1/n)$-close to S needs nonnegative rank $2^{\Omega(n)}$.

Edmonds gave a complete description of the facets of the perfect matching polytope for the complete n-vertex graph K_n [Edm65]. The key are the *odd-set* inequalities: for a perfect matching M, viewed as a vector $M \in \{0,1\}^{\binom{n}{2}}$ of weight $m = n/2$, and an odd-sized set $U \subseteq [n]$, the associated inequality says $|\delta(U) \cap M| \geq 1$, where $\delta(U) \in \{0,1\}^{\binom{n}{2}}$ denotes the cut induced by U. In addition, there are $O(n^2)$ degree

and nonnegativity constraints. Thus the corresponding slack matrix S has columns indexed by all perfect matchings M in K_n and rows indexed by odd-sized sets U with entries $S_{UM} = |\delta(U) \cap M| - 1$. There are $O(n^2)$ additional rows for the degree and nonnegativity constraints.

In Theorem 20 in the appendix, we show that the m-bit function $g(z) = |z| - 1$ can be approximated (in expectation) up to exponentially small error with quantum query complexity $O(m^{1/2+\varepsilon} \log m)$. Define $f_M(x) = g(x_M)$, where x_M denotes the restriction of n-bit string x to the m positions in the support of M. Applying Theorem 18 and adding $O(n^2)$ rows to account for the other constraints gives:

Theorem 19 *For every $\varepsilon > 0$ there exists a matrix $\tilde{S}$ of psd rank $2^{O(n^{1/2+\varepsilon}(\log n)^2)}$ such that*

1. *$S_{UM} - 2^{-(n/2)^{2\varepsilon}} \leq \tilde{S}_{UM} \leq S_{UM}$ for the UM-entries where $|\delta(U) \cap M| > (n/2)^{2\varepsilon}$;*
2. *$\tilde{S}_{xy} = S_{xy}$ for all other entries.*

Acknowledgments. We thank Srinivasan Arunachalam, David Steurer, Mario Szegedy and Henry Yuen for useful discussions, Sebastian Pokutta for useful discussions and for pointing us to [BP15], and James Lee for sending us a version of [LRS14].

References

[AYW14] S. Arunachalam, H. Yuen, and R. de Wolf. Unpublished manuscript, August 2014.

[BBC+01] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4):778–797, 2001. Earlier version in FOCS'98.

[BCW98] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proceedings of 30th ACM STOC*, pages 63–68, 1998. quant-ph/9802040.

[BCWZ99] H. Buhrman, R. Cleve, R. de Wolf, and Ch. Zalka. Bounds for small-error and zero-error quantum algorithms. In *Proceedings of 40th IEEE FOCS*, pages 358–368, 1999. cs.CC/9904019.

[BGP14] G. Blekherman, J. Gouveia, and J. Pfeiffer. Sums of squares on the hypercube. arXiv/1402.4199, 18 Feb 2014.

[BHMT02] G. Brassard, P. Høyer, M. Mosca, and A. Tapp. Quantum amplitude amplification and estimation. In *Quantum Computation and Quantum Information: A Millennium Volume*, volume 305 of *AMS Contemporary Mathematics Series*, pages 53–74. 2002. quant-ph/0005055.

[BP15] G. Braun and S. Pokutta. The matching polytope does not admit fully-polynomial size relaxation schemes. In *To appear in Proceedings of SODA*, 2015.

[BW01] H. Buhrman and R. de Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of 16th IEEE Complexity (CCC)*, pages 120–130, 2001.

[BW02] H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: A survey. *Theoretical Computer Science*, 288(1):21–43, 2002.

[CLRS13] S. O. Chan, J. R. Lee, P. Raghavendra, and D. Steurer. Approximate constraint satisfaction requires large LP relaxations. In *Proceedings of 54th IEEE FOCS*, pages 350–359, 2013.

[Edm65] J. Edmonds. Maximum matching and a polyhedron with 0,1-vertices. *Journal of research of the National Bureau of Standards–B*, 69B(1,2):125–130, 1965.

[FMP+12] S. Fiorini, S. Massar, S. Pokutta, H. R. Tiwary, and R. de Wolf. Linear vs. semidefinite extended formulations: Exponential separation and strong lower bounds. In *Proceedings of 44th ACM STOC*, pages 95–106, 2012.

[FSP13] H. Fawzi, J. Saunderson, and P. Parrilo. Equivariant semidefinite lifts and sum-of-squares hierarchies. arXiv:1312.6662, Dec 23, 2013.

[GPT13] J. Gouveia, P. Parrilo, and R. Thomas. Lifts of convex sets and cone factorizations. *Mathematics of Operations Research*, 38(2):248–264, 2013. arXiv:1111.3164.

[Gri01] D. Grigoriev. Complexity of Positivstellensatz proofs for the knapsack. *Computational Complexity*, 10:139–154, 2001.

[Gro96] L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of 28th ACM STOC*, pages 212–219, 1996. quant-ph/9605043.

[KN97] E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, Cambridge, 1997.

[Kre95] I. Kremer. Quantum communication. Master's thesis, Hebrew University, Computer Science Department, 1995.

[Las01] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[Lau03] M. Laurent. Lower bound for the number of iterations in semidefinite hierarchies for the cut polytope. *Mathematics of operations research*, 28(4):871–883, 2003.

[LRS14] J. R. Lee, P. Raghavendra, and D. Steurer. Lower bounds on the size of semidefinite programming relaxations. arXiv:1411.6317, Nov 24, 2014.

[LRST14] J. R. Lee, P. Raghavendra, D. Steurer, and N. Tan. On the power of symmetric LP and SDP relaxations. In *Proceedings of 29th IEEE Complexity (CCC)*, pages 13–21, 2014.

[Mid04] G. Midrijanis. Exact quantum query complexity for total Boolean functions. quant-ph/0403168, 23 Mar 2004.

[MP87] M. Minsky and S. Papert. *Perceptrons*. MIT Press, 1987.

[NC00] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[Pad89] M. Padberg. The boolean quadric polytope. *Mathematical programming*, 45:139–172, 1989.

[Par00] P. Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, California Institute of Technology, 2000.

[Rot14] T. Rothvoß. The matching polytope has exponential extension complexity. In *Proceedings of 46th ACM STOC*, pages 263–272, 2014.

[SA90] H. D. Sherali and W. P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming. *SIAM Journal on Discrete Mathematics*, 3:411–430, 1990.

[Sho87] N. Z. Shor. An approach to obtaining global extremums in polynomial mathematical programming problems. *Cybernetics*, 23:695–700, 1987.

[Swa86] T. Swart. P = NP. Technical report, University of Guelph, 1986. Revision 1987.

[Wol02] R. de Wolf. Quantum communication and complexity. *Theoretical Computer Science*, 287(1):337–353, 2002.

[Wol03] R. de Wolf. Nondeterministic quantum query and quantum communication complexities. *SIAM Journal on Computing*, 32(3):681–699, 2003.

[Yan91] M. Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3):441–466, 1991. Earlier version in STOC'88.

[Yao93] A. C-C. Yao. Quantum circuit complexity. In *Proceedings of 34th IEEE FOCS*, pages 352–360, 1993.

[YFGT12] S. Fiorini Y. Faenza, R. Grappe, and H. R. Tiwary. Extended formulations, nonnegative factorizations, and randomized communication protocols. In *Proceedings of ISCO'12*, pages 129–140, 2012.

A A tailored quantum search algorithm

The *search problem* is the following: we have an m-bit input z that we can access by means of queries, and our goal is to find an index $i \in [m]$ such that $z_i = 1$. Such an i will be called a "solution". The number of solutions is the Hamming weight of the input, denoted $|z|$. Grover's algorithm [Gro96, BHMT02] solves this problem using $O(\sqrt{m})$ queries. We will use the following two variants:

- There is a quantum algorithm using $O(\sqrt{m/t})$ queries that finds a solution with probability at least 1/2 if $|z| \in [t, 2t]$.
- There is a quantum algorithm using $O(\sqrt{m/t})$ queries that finds a solution with certainty if $|z| = t$.

We combine these variants of Grover to prove the following theorem, similar to [BCWZ99, Theorem 3]:

Theorem 20 *For every integer $\ell > 0$ there exists a quantum algorithm that makes $O(\sqrt{m\ell}\log m)$ queries to input $z \in \{0,1\}^m$ and that has the following properties:*

1. *If $z = 0^m$ then the algorithm outputs "no solution" with certainty.*
2. *If $|z| \in \{1, \ldots, \ell\}$ then the algorithm outputs a solution with certainty.*
3. *If $|z| > \ell$ then the algorithm outputs a solution with probability $\geq 1 - 2^{-\sqrt{\ell |z|}}$.*

Proof: The algorithm is as follows:

1. Run exact Grover ℓ times, once for each of the possibilities $t = 1, 2, \ldots, \ell$.
2. For $i = \lfloor \log \ell \rfloor, \ldots, \lfloor \log m \rfloor$: Run $\left\lceil \sqrt{\ell 2^{i+1}} \right\rceil$ times the version of Grover that assumes $|z| \in [2^i, 2^{i+1}]$.
3. Check each of the indices produced by these runs (using one query per index).
4. Output a solution if one was found, and output "no solution" otherwise.

Clearly, the algorithm behaves as promised if $|z| \leq \ell$. Now suppose $|z| > \ell$ and let i be the unique integer such that $|z| \in [2^i, 2^{i+1})$. For that i, each of the $\left\lceil \ell\sqrt{2^{i+1}} \right\rceil$ runs of Grover has probability $\geq 1/2$ of producing a solution, hence the probability of *not* finding a solution is $\leq 2^{-\lceil \sqrt{\ell 2^{i+1}} \rceil} \leq 2^{-\sqrt{\ell |z|}}$ in this case.

It remains to bound the query complexity of the algorithm. The number of queries used in step 1 is

$$\sum_{t=1}^{\ell} O(\sqrt{m/t}) = O(\sqrt{m\ell}).$$

The number of queries used in step 2 is

$$\sum_{i=\lfloor \log \ell \rfloor}^{\lfloor \log m \rfloor} \left\lceil \sqrt{\ell 2^{i+1}} \right\rceil O\left(\sqrt{m/2^i}\right) = O(\sqrt{m\ell} \log m).$$

The total number of runs of (versions of) Grover's algorithm is

$$\ell + \sum_{i=\lfloor \log \ell \rfloor}^{\lfloor \log m \rfloor} \left\lceil \sqrt{\ell 2^i} \right\rceil = O(\sqrt{m\ell}).$$

Since each such run produces one index that needs to be checked, the number of queries made in step 3 is $O(\sqrt{m\ell})$. Thus the overall query complexity is $O(\sqrt{m\ell} \log m)$ as promised. □

We can derive from this a function $f : \{0,1\}^m \to \mathbb{R}_+$ that approximates $|z| - 1$ extremely well, and that has quantum query complexity in expectation not much bigger than $\sqrt{m}$:

Theorem 21 *For every $\varepsilon > 0$ there exists a function $f : \{0,1\}^m \to \mathbb{R}_+$ satisfying $\mathrm{QE}(f) = O(m^{1/2+\varepsilon} \log m)$ and*

1. *$f(0^m) = 0$.*
2. *If $|z| \in \{1, \ldots, \ell\}$ then $f(z) = |z| - 1$.*
3. *If $|z| > \ell$ then $|z| - 1 - 2^{-m^{2\varepsilon}} \leq f(z) \leq |z| - 1$.*

Proof: Set $\ell = m^{2\varepsilon}$. Run the algorithm of Theorem 20, which uses $O(m^{1/2+\varepsilon} \log m)$ queries. If $|z| \geq 1$, it finds a solution (i.e., an $i \in [m]$ such that $z_i = 1$) with very high probability. If it did not find a solution the algorithms outputs 0. If, on the other hand, i is a solution then the algorithm queries a uniformly random index $j \neq i$ and outputs $z_j \cdot (m-1)$. Let $f(z)$ be the expected output of this algorithm on input z.

If $z = 0^m$ the algorithm always outputs 0, establishing the first property. If $|z| \in \{1, \ldots, \ell\}$ then i is a solution with certainty, and the expected value of the output is $\Pr[z_j = 1] \cdot (m-1) = \frac{|z|-1}{m-1} \cdot (m-1) = |z| - 1$, establishing the second property. If $|z| > \ell$ then the algorithm finds a solution except with probability $2^{-m^{2\varepsilon}}$, which implies the third property. □